A suite of *de novo c*-type cytochromes for functional oxidoreductase engineering.

Watkins, Daniel W.[a]; Armstrong, Craig T.[a]; Beesley, Joseph L.[a,b]; Marsh, Jane E.[a]; Jenkins, Jonathan [a]; Sessions, Richard B.[a]; Mann, Stephen[b]; Anderson, J. L. Ross[a].

[a]School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, UK
[b]School of Chemistry, University of Bristol, Bristol BS8 1TS, UK
corresponding author - ross.anderson@bristol.ac.uk; tel: +44 (0)117 331 2151

Keywords

Maquette, heme C, cytochrome *c*, protein design, four-helix bundle, oxygen binding, oxygen activation, *E. coli* cytochrome *c* biogenesis system I, enzyme design

Abbreviations

CTM – *c*-type maquette; CBM – consensus heme C binding motif; Ccm - *E. coli* cytochrome *c* biogenesis system I; cCTM – di-heme C binding maquette; bCTM – mixed heme B/heme C binding maquette; CD – circular dichroism; MD – molecular dynamics simulation; TEV – tobacco etch virus; HPLC – reversed-phase high-pressure liquid chromatography; OTTLE – optically-transparent thin-layer electrochemistry.

Abstract

Central to the design of an efficient *de novo* enzyme is a robust yet mutable protein scaffold. The maquette approach to protein design offers precisely this, employing simple four-α-helix bundle scaffolds devoid of evolutionary complexity and with proven tolerance towards iterative protein engineering. We recently described the design of C2, a *de novo* designed *c*-type cytochrome maquette that undergoes post-translational modification in *E. coli* to covalently graft heme onto the protein backbone *in vivo*. This *de novo* cytochrome is capable of reversible oxygen binding, an obligate step in the catalytic cycle of many oxygen-activating oxidoreductases. Here we demonstrate the flexibility of both the maquette platform and the post-translational machinery of *E. coli* by creating a suite of functional *de novo* designed *c*-type cytochromes. We explore the engineering tolerances of the maquette by selecting alternative binding sites for heme C attachment and creating di-heme maquettes either by appending an additional heme C binding motif to the maquette scaffold or by binding heme B through simple bis-histidine ligation to a second binding site. The new designs retain the essential properties of the parent design but with significant improvements in structural stability. Molecular dynamics simulations aid the rationalization of these functional improvements while providing insight into the rules for engineering heme C binding sites in future iterations. This versatile, functional suite of *de novo c*-type cytochromes shows significant promise in providing robust platforms for the future engineering of *de novo* oxygen-activating oxidoreductases.

Introduction

Natural cytochromes are an exceptionally diverse group of heme-containing proteins that exhibit an impressive breadth of function, ranging from simple interprotein electron transfer reactions (e.g. cytochrome *c* [1]) to sophisticated multistep catalysis (e.g. cytochrome P450 [2]) in which molecular oxygen can be stereo- and regiospecifically inserted into relatively unreactive organic molecules [3,4]. Despite the complexity evident in many natural cytochromes, fairly simple engineering rules are known to underpin some of their key functions and reactivities [5], and there has been significant recent success in imprinting these features onto simple *de novo* protein scaffolds, recapitulating functions and catalytic intermediates integral to cytochrome P450 and heme-containing dioxygenase activity [6–8]. The ability to access such activity within *de novo* protein scaffolds is particularly attractive from an industrial perspective [9,10] and will test our fundamental understanding of oxidoreductase activity while ultimately facilitating the construction of bespoke enzymes tailored for specific applications.

The Maquette approach [11–14] has been used to great effect in the construction of functional, *de novo* proteins that exhibit sophisticated features common to natural cytochromes and

oxidoreductases. This non-computational design strategy involves the iterative incorporation of functional engineering elements into simple, robust four-helix bundle proteins that encode only the minimum sequence information required for secondary and tertiary structure formation, solubility and cofactor binding [13]. These proteins contain no natural protein sequences and, due to the absence of unnecessary evolutionary history and complexity, are consequently tolerant of extensive modification without disruption to the protein structure or function [11]. Thus, these maquettes provide ideal scaffolds for a tractable enzyme design process. To date, maquettes have been designed that bind a diverse array of cofactors including hemes [15], chlorins [16], flavins [17] and quinones [18], replicating features of natural cytochromes and oxidoreductases such as intermolecular electron transfer [7], proton-coupled electron transfer [19] and reversible oxygen binding [6]. Tetrapyrrole binding sites in these maquettes are typically composed of two histidine residues placed in appropriate core positions on opposing helices, with the imidazole moieties of the histidine residues axially coordinating the central metal ion in the tetrapyrrole [12]. A somewhat fluid, mobile hydrophobic protein core likely aids tetrapyrrole entry [20] and numerous hydrophobic interactions between bound tetrapyrrole and hydrophobic sidechains such as phenylalanine are formed [21]. Despite the success in producing functional maquettes that bind heme B in this manner, there are significant limitations to this method: the heme B ligation scheme has been restricted to bis-histidine coordination; the orientation of the heme is uncontrolled within the binding site; although the maquettes are expressed *in vivo*, levels of spontaneous heme B incorporation in *E. coli* are typically low and heme B is added *in vitro* [15,21].

We recently reported the construction of an artificial, *in vivo* assembled heme C containing maquette (CTM), described in this paper as ‘C2’ [8]. Though structurally similar to heme B, heme C differs by virtue of its connection to the protein. In addition to hydrophobic contacts with core hydrophobic side chains and electrostatic interactions between heme propionates and positively charged amino acid sidechains, heme B is non-covalently linked to the protein *via* axial heme ligation from metal-coordinating amino acid sidechains (e.g. histidine, methionine) [22]. Heme C maintains these hydrophobic, electrostatic and metal-coordinating links to the protein, but is also covalently attached to cysteine sidechains *via* thioether linkages to the periphery of the heme macrocycle [23]. This post-translational modification is performed in Gram-negative bacteria by the *c*-type cytochrome maturation apparatus (known as the cytochrome *c* biogenesis system I, Ccm, in *E. coli*) within the periplasmic envelope [23]. A heme C recognition motif (usually CXXCH) is essential to achieving this modification in *E. coli*, with the cysteine sidechains in the motif providing the covalent anchoring sites for thioether bond formation and the histidine acting as the proximal coordinating ligand for the heme iron [24]. This covalent modification allows for a site- and orientation-specific heme attachment within

the maquette, increasing the degree of engineering control available to the designer [8]. It also raises the unique possibility of studying the activity of a functional *de novo* protein or enzyme within a living organism, and subsequently replacing natural cellular machinery with designed, artificial cytochromes.

C2 [8] is a 138 residue, antiparallel 4-helix bundle with nearly identical α-helices and two heme binding sites in the protein core: one between the first and third helices (from the N-terminal) and the other between the second and fourth helices (Fig.1). Each helix consists of binary patterned hydrophobic and hydrophilic residues designed to drive protein folding by shielding hydrophobic amino acid side chains from water in a hydrophobic core. It has a helix-stabilizing N-terminal cap and simple glycine/serine loops link the helices in the selected topology. A CIACH Ccm recognition motif is situated on the second helix and following Sec translocon-mediated transport to the *E. coli* periplasm, the protein is post-translationally modified by the Ccm to append heme *in vivo*. A second histidine residue situated on the fourth helix provides the distal axial ligand at the heme iron. An additional bis-histidine tetrapyrrole-binding site is situated between helices one and three and may be equipped with heme B or other tetrapyrroles *in vitro*. When loaded with heme B in the second tetrapyrrole-binding site, C2 has two electrochemically distinct redox active sites, thus producing a nascent electron transfer chain for defined vectorial electron transfer.

When exposed to molecular oxygen, ferrous C2 undergoes rapid ligand exchange at the distal histidine [8], forming an oxyferrous species in a manner analogous to the hexacoordinate O_2 binding globins such as neuroglobin [25]. Since oxyferrous heme is an obligate intermediate in oxygen-activating oxidoreductases such as the cytochromes P450 [2], there is a strong motivation to further engineer maquettes such as C2 to recapitulate other reactive intermediates in the catalytic cycles of these powerful hydroxylation enzymes and look towards the design of heme-containing *de novo* monooxygenases [9,10].

To maximize our chances of achieving this ambitious goal, we require not only a single protein as a starting proto-oxidoreductase scaffold but a range of suitable and robust *de novo* proteins for further design. In addition, C2 has relatively low thermal stability when the second tetrapyrrole binding site is unoccupied [8] and we wished to address this issue while probing the relationship between heme placement within the maquette and the resulting biophysical properties of the CTMs. Here we describe the design, expression, purification and *in vitro/silico* characterisation of a suite of versatile, engineering tolerant heme-binding maquettes created simply by shifting the heme C binding motif (CBM) to the other locations within the original heme B binding sites. These new proteins bind heme

C *via* these alternate CBMs in C2, creating 3 new mono-heme CTMs that can be loaded with heme B *in vitro* to form mixed heme B/C proteins (bCTMs) and 4 new di-heme C proteins (cCTMs). The mono-heme CTMs retain the essential functions of C2, but exhibit some differences in biophysical properties such as oxygen binding kinetics, heme midpoint potential and thermal stability, thus allowing the designer to select a CTM scaffold with properties fine-tuned for a specific application.

Materials and Methods

Expression vectors and mutagenesis – C2 was previously cloned [8] into a modified pMal-p4x (NEB) expression vector with the maltose binding protein (MBP) sequence deleted, leaving the MBP Sec translocon recognition sequence (MKIKTGARILALSALTTMDFSASALAK) thus enabling periplasmic translocation of the artificial c-type cytochrome. A His_6 tag was included to facilitate purification by immobilised metal affinity chromatography (IMAC) and a Tobacco Etch Virus (TEV) recognition sequence (ENLYFQ/G) was added between the His_6 tag and the protein sequence to allow the removal of the purification tag. For mutagenesis, a site-directed ligase-independent mutagenesis protocol was followed [26]. Specifically, this involved mutating CBM2 from CIAC to IWKQ and CBM1, 3 and 4 from IWKQ to CIAC. The operon for *c*-type cytochrome maturation apparatus (*CcmA-H*) was previously cloned into a constitutive expression vector with chloramphenicol resistance, creating the vector known as pEC86 [27].

Cell strains and expression – CTMs were co-transformed with pEC86 into *E. coli* T7 Express (NEB) competent cells. To ensure the presence of both plasmids in the *E. coli*, carbenicillin and chloramphenicol were added to media and plates at concentrations of 50 µg/L and 34 µg/L respectively. A starter culture was prepared with 50 ml lysogeny-broth (LB) and left to grow for 16 hours at 37°C and shaking at 200 RPM. This was used to inoculate 1 L LB and upon reaching an OD_{600nm} of 0.6, CTM expression was induced with 1 mM IPTG. After four hours of expression, cells were harvested by centrifugation at 4000 xg for 20 minutes.

Protein purification – All artificial c-type cytochromes were purified from *E. coli,* as previously described [8]. 1 mM phenylmethanesulfonyl fluoride was added to cell pastes. Cells were then placed on ice, lysed by sonication (Soniprep 150 plus MSE) and subsequently centrifuged for 40000 xg for 30 minutes. Cell lysate was decanted for purification by nickel affinity chromatography.

A 5 mL HisTrap HP IMAC column (GE Healthcare) connected to an ÄKTA start purification system (GE Healthcare) was equilibrated in lysis buffer (40 mM imidazole, 300 mM NaCl, 50 mM K

Phosphate pH 8.0). Cell lysate was filtered (0.22 µM syringe filter, Millipore) and loaded onto the column at 5 mL/min *via* the sample loading valve. Upon completion, the column was equilibrated with 50 mL of lysis buffer and then a linear gradient of 0-100% elution buffer (250 mM imidazole, 300 mM NaCl, 50 mM K Phosphate pH 8.0) was applied at 2 mL/min for 15 minutes. 1 mL fractions were collected and analysed by SDS PAGE. Fractions containing maquettes were transferred into a 3 kDa semi-permeable membrane and dialysed for 16 hours in a 5 L solution of 0.5 mM EDTA and 20 mM Tris pH 8.0.

Following dialysis, DTT (1 mM) and TEV protease (1 µM) was added to remove the hexahistidine tag from the N-terminus of maquettes. Samples were mixed and left at room temperature for five hours. Samples were then centrifuged (4000 xg, 10 minutes) and filtered (0.22 µM syringe filter, Millipore). The ÄKTA start purification system was set up as described above. Cleaved maquettes eluted in the column flow-through and were transferred to a 10 kDa Vivaspin semi-permeable centrifugal concentrator (Sartorius Stedim). Samples were centrifuged (4000 xg) until sample volumes were approximately 1.5 mL.

A Varian-SD high pressure liquid chromatography system connected to a C-18 reversed-phase HPLC column (Phenomenex) was equilibrated in 30% acetonitrile and 0.1% trifluoroacetic acid at 5 mL/min. 0.5 mL samples were loaded *via* the injection valve. Elution absorbance was monitored at 280 and 395 nm. After 5 minutes, a 30-55% linear gradient of acetonitrile was applied. Holo-maquette was manually collected into a single fraction. All proteins were freeze-dried to remove acetonitrile after HPLC and resuspended in 100 mM KCl, 20 mM K phosphate, pH 7.5. These buffer conditions were used for all analyses unless stated otherwise.

UV-visible spectroscopy – An Agilent Cary-60 UV-visible spectrophotometer was used to record all UV-visible spectra. Typically, CTM concentration when recording spectra was 5 µM. Ferrous CTM spectra were recorded after adding a few grains of sodium dithionite. Pyridine hemochrome spectra were measured to confirm the association of heme C to the protein backbone by preparing CTM solutions as described in Berry *et. al.* [28]. CTM concentrations were calculated from the pyridine hemichrome and hemochrome ε_{550nm} (32,700 $M^{-1}cm^{-1}$ and 8,430 $M^{-1}cm^{-1}$, respectively), assuming a stoichiometric ratio of heme C to protein of 1:1 for CTMs and 2:1 for cCTMs.

Heme titrations – To assess CTM heme B binding, 0.5 µM aliquots of hemin in DMSO were titrated into a 5 µM solution of CTM and a UV-visible spectrum measured after equilibration. The Absorbance 409 nm was then plotted against hemin concentration to determine the heme binding stoichiometry.

Optically-transparent thin-layer electrochemistry (OTTLE) potentiometry – heme redox chemistry was analysed with OTTLE potentiometry. CTMs were exchanged into 500 mM KCl, 10% glycerol, 100 mM K phosphate, pH 7.5. The following redox mediators (with midpoint transition potentials quoted vs Nernst hydrogen electrode (E_m)) were added to ensure rapid equilibration: 20 µM benzyl viologen (-311 mV), 20 µM anthroquinone-2-sulfonate (-225 mV), 20 µM phenazine (-180 mV), 25 µM 2-hydroxy-1,4-napthoquinone (-152 mV), 6 µM indigotrisulfonate (-90 mV) and 50 µM duroquinone (5 mV). A platinum gauze working electrode was submerged in 50-100 µM CTM in a homemade OTTLE cell constructed from a flat EPR quartz cuvette (Wilmad, USA). A platinum counter electrode and a Ag/AgCl reference electrode (BASi, USA) calibrated against bovine cytochrome *c* were submerged in a layer of the previously described buffer without glycerol, carefully (to prevent agitation) pipetted on top of 80 µL of CTM solution. Potentials were applied across the cell using a Biologic SP-150 potentiostat and potentials were applied at 25°C from 50 to -350 mV vs NHE, then cycled back to 50 mV vs NHE to ensure that hysteresis did not occur. Heme soret λ_{max} was plotted vs applied potential and data were fitted to a Nernst fitting functions adapted for mono- (Eq.1) or di- (Eq.2) heme maquettes.

Eq.1: $f(x) = (A+B*10^{((E_m-x)/59)})/(1+10^{((E_m-x)/59)})$

Eq.2: $f(x) = (A*(10^{((x-E_m1)/59)})+B+C*(10^{((E_m2-x)/59)}))/(1+10^{((x-E_m1)/59)}+10^{((E_m2-x)/59)})$

'A' and 'B' are y-axis values at 100% oxidised and reduced CTM, respectively; 'C' reflects the absorbance contribution of each redox species for a di-heme-maquette; 'E_m', 'E_m1' and 'E_m2' are the heme midpoint potentials.

Circular dichroism spectroscopy (CD) – CD was performed with a JASCO J-815 CD polarimeter. Protein concentration was between 0.015-0.15 mg/ml (1-10 µM) and buffer conditions were 100 mM KCl, 20 mM K phosphate pH 7.5. Samples were loaded into a 1 mm pathlength CD quartz cuvette. Far-UV CD spectra parameters were 100 nm/min scanning rate and 50 mdeg sensitivity. For thermal denaturation traces a ramp rate of 40°C/hour with 1°C intervals was applied and 222 nm ellipticity was measured at each temperature. Raw elliptical units were converted to mean residue ellipticity (MRE) for comparison by normalising to peptide bond concentration and cuvette pathlength. Thermal denaturation midpoints (T_ms) were determined from the x-axis intercept of the second derivative of thermal denaturation traces.

Molecular dynamics simulations (MD) – Starting MD models were based on the design of C2. First, helices were built and placed in an antiparallel topological conformation. An N-cap and loops were added and heme C built in with heme vinyl chains placed in 0.7 Å proximity of corresponding cysteine

sulfur groups. During pre-simulation set up, thioether bonds were automatically generated between heme vinyl chains and cysteine side chains. The remaining CTMs and cCTMs were built by modifying the C2 starting model. The CHARMM27 forcefield [29] (contains heme MD parameters) was used in conjunction with GROMACS (4.5.5 and above) MD software [30]. MD simulations were performed under periodic boundary conditions with PME long range electrostatics and an NPT ensemble, maintaining the temperature and pressure at 300 K and 1 bar respectively. The periodic box was filled with TIP3P water and 150 mM NaCl with protein groups ionised consistent with pH 7.5. Structures were saved every 0.1 ns for further analysis. GROMACS tools were used to calculate Cα RMSD traces and other properties.

Measuring CTM oxygen-binding kinetics – 5 μM CTM was deoxygenated in an anaerobic glove box under N_2 ($[O_2]$ < 5 ppm; Belle Technology) overnight. Buffer conditions were 100 mM KCl, 20 mM K phosphate pH 7.5. Prior to mixing, CTMs were reduced with a stoichiometric quantity of sodium dithionite. Using an SX20 Stopped Flow Spectrophotometer (Applied Photophysics) a single-mixing stopped-flow experiment was conducted at 15°C in the anaerobic glove box. 5 μM CTM and buffer containing 540 μM of oxygen as measured using an oxygen probe (Lutron PDO-520) were mixed and UV-visible spectra were measured for 60 seconds using a diode array detector. Data were analysed by plotting oxyferrous heme absorbance at 568 nm vs time, and fitting to two consecutive single exponential functions: the first between 0 and 0.1 seconds and the second from 0.1 to 60 seconds from which oxygen binding and autoxidation rates were calculated.

Results

Protein design – To construct new heme C maquettes, CBM2 was first converted to a heme B binding site by mutating CIAC to IWKQ (Fig.1A). IWKQ was then replaced with CIAC at the CBM1, CBM3 or CBM4 sites to produce the three new mono-heme C binding proteins (CTMs) named C1, C3 and C4 (Fig.1B). These new CTMs retain a heme B binding site and are therefore capable of forming mixed heme B/heme C maquettes (bCTMs). Each heme-binding site is composed of parallel alpha helices; therefore, for C1 and C3 or C2 and C4 the orientation of heme C is inverted, leaving the heme propionates pointing into either the protein core (C1 and C2) or bulk solvent (C3 and C4) (Fig.1C). In addition to the monoheme CTMs four di-heme C maquettes (cCTMs) - C1C2, C1C4, C2C3 and C3C4 - were constructed by converting a CBM in the other tetrapyrrole binding site to a heme C binding motif (Fig.1B). This suite of proteins provides a test bed system for exploring the key relationships between heme location and CTM biophysical properties. These include assessing the effects of swapping heme orientation within the same binding site, probing the heme C binding

capacity of the CTMs and assessing the engineering tolerances of the CTMs to gross changes enforced by the CBM movement.

Expression and purification – CTM vectors were co-transformed into *E. coli* with the pEC86 plasmid containing Ccm genes for post-translational heme C incorporation. Proteins were expressed for four hours and harvested by centrifugation. With the exception of C3C4, the pelleted cells were noticeably red, indicating both successful CTM expression and efficient post-translational heme C incorporation (Fig.1D). Cell lysates were loaded onto a nickel affinity chromatography column where the CTMs were observed binding at the top of the column (*via* the His_6 tag) as evidenced by the accumulation of red colour. A linear gradient of imidazole was applied to elute the proteins and a large 280 nm peak eluted at approximately 80 mM imidazole, corresponding to red CTM fractions (SiFig.1). SDS-PAGE confirmed the presence of CTMs with an intense band visible at approximately 17 kDa (apo-CTM + His_6 tag = 17176 Da; apo-cCTM + His_6 tag = 17029 Da; CTM + His_6 tag = 17861 Da; cCTM + His_6 tag = 17714 Da) (SiFig.1). The elution peak integral for C1 and C1C4 was large relative to the other variants, whereas for C3C4 there was no noticeable elution peak or SDS PAGE band, indicating the absence of soluble protein. Taken with the lack of red colour in the C3C4 cell pellet, it is most likely that C3C4 does not undergo the post-translational heme incorporation and insoluble protein is instead produced. We were therefore unable to purify and characterise apo- or holo-C3C4.

CTMs were then treated with TEV protease to remove the His_6 tag. An expected decrease in mass (2052 Da) occurred upon removal of the His_6 tag by TEV protease, as visualised by SDS PAGE (SiFig.2). Protein solutions were then reloaded onto a nickel affinity column and the cleaved CTMs were collected from the column flow through. (SiFig.3). There was a pronounced variation in the efficiency of TEV cleavage across the CTMs, with approximately 70-90% of CTMs successfully cleaved by TEV protease as estimated from the ratio of 280 nm peak integrals for the column flow through (cleaved protein) compared to the column elution (uncleaved protein).

Following TEV cleavage and the second IMAC step, CTMs were purified by HPLC chromatography to remove maquettes lacking heme C. A small 280 nm non-binding injection peak was confirmed as protein impurities and imidazole (SiFig.4). A gradient of 30-55% acetonitrile was applied over ten minutes at 3 ml/min with CTM typically eluting after 7 minutes, indicative from an increase in 280 nm and 395 nm absorbance corresponding to protein and heme absorption respectively. Interestingly, an additional small peak eluted containing heme and was determined to be solely heme B, suggesting a small amount of purified CTM was loaded with heme B *in vivo*. Also, in the CTM traces at 280 nm a small peak eluted just prior to the main CTM elution peak, which was subsequently determined by

matrix-assisted laser desorption/ionisation time of flight spectrometry (MALDI-TOF) to be apo-CTM (Mw = 15124 Da). This suggests a small fraction of the expressed CTM proteins were not modified by the *E. coli* Ccm to include heme C. This apoprotein was also evident in the cCTM chromatograms (Apo cCTM Mw = 14977 Da) and a further peak indicated that a small fraction of cCTMs contained only one heme C (Mw = 15615 Da). For C1C2 no di-heme C protein was produced, suggesting the Ccm was not capable of fully processing C1C2. MALDI-TOF mass spectra confirmed the remaining CTM and cCTM masses indicating that they contained one or two heme Cs as designed.

UV-visible spectroscopic characterisation – The UV-visible absorption spectra of the monoheme CTMs, and the diheme C1C4 and C2C3 were recorded to compare their spectroscopic properties (Fig.2A,B). The UV-visible spectra of new designs were near identical to C2 [8], with a ferric soret λ_{max} of 409 nm and a ferrous Soret, alpha and beta λ_{max} of 418 nm, 552 nm and 529 nm respectively. C2 was previously confirmed to have a bis-histidine heme iron ligation scheme. The similarities in UV-visible absorption properties suggest that this is also the case for C1, C3 and C4. The UV-visible spectrum of C2C3 exhibited significant scattering indicating some degree of protein aggregation. Upon closer inspection of C2C3, small aggregates were visible by eye, suggesting the protein was not stable in solution. This left C1C4 as the only solution stable, holo-cCTM with good expression levels that was efficiently processed by the Ccm.

Pyridine hemichrome and hemochrome solutions of CTMs were prepared using the protocol described by Berry *et. al.* [28]. In all cases, the hemochrome spectra exhibited an alpha band λ_{max} at 550 nm, indicative of heme C absorption and confirming the heme type attached to the CTM scaffolds (Fig.2C). Heme concentration was calculated from the difference in pyridine hemochrome and hemichrome absorbance with the ε_{550nm}.

CTM heme B binding properties – Heme titrations were conducted with CTMs to assess heme B binding and thus the production of mixed heme maquettes, bCTMs. It was previously demonstrated that C2 binds heme B in its second tetrapyrrole binding site with high affinity [8]. To probe the stoichiometry of heme B binding, hemin was titrated into the CTMs and a sharp, linear increase in soret λ_{max} absorbance prior to saturation of the binding site was observed. After the saturation point was reached, a smaller increase in absorbance was observed (Fig.3). The heme B binding data confirm that the new CTMs exhibit similar heme B binding properties to C2, all displaying near 1:1 ratios of heme B:CTM thus demonstrating that in all cases the second tetrapyrrole binding site remains functional regardless of CBM position.

OTTLE potentiometry analysis of heme electrochemical properties – Heme C in C2 was previously determined to exhibit reversible redox properties with a midpoint potential vs Nernst hydrogen electrode (E_m) of -198 mV [8]. CTM midpoint potentials were determined using optically-transparent thin layer electrochemisty (OTTLE) potentiometry (Fig.4). Data were fitted to Nernst functions adapted to mono- (Eq.1) and di- (Eq.2) heme binding proteins. Monoheme CTMs exhibit midpoint potentials spanning almost 40 mV, from -220 mV (C4) to -181 mV (C3) (Fig.4A-C, Tab.1). The midpoint potentials were also recorded for the mixed diheme bCTMs (Fig.4D-F). The heme B of BC2 was previously shown to have a 50 mV more negative midpoint potential than heme C, deduced from the differing UV-visible properties of heme B and heme C. The new bCTMs were also determined to contain electrochemically distinct redox sites, also with differences in E_m between heme B and heme C of approximately 50 mV (Fig.4, Tab.1). Interestingly, loading heme B into C4 had a dramatic effect on the heme C E_m with a shift of +44 mV. The diheme C1C4 was determined to have two distinct heme C midpoint potentials of -207 mV and -154 mV (Fig.4G) again leading to a maquette with a greater than 50 mV difference in heme midpoint potentials, though in this case identical hemes occupy the two binding sites. Since both heme Cs in C1C4 share almost identical spectroscopic properties, we were unable to assign potentials to specific CBMs, though we would surmise that since C4 undergoes a significant shift in potential with heme B present in BC4, then the higher potential heme in C1C4 is likely to occupy the C4 site.

Structural characterization with Circular dichroism (CD) and molecular dynamics (MD) – Circular dichroism spectroscopy was used to probe and quantify maquette secondary structure, measure thermal stability and to rule out concentration-dependent oligomerisation. Far-UV CD spectra were recorded for 5 μM and 0.5 μM CTM solutions at a range of temperatures from 5°C to 85°C (Fig.5A,B). Helical secondary structure was dominant at 25°C, as evidenced by the signature spectra of α-helices with a negative elliptical λ_{max} at 222 nm. No differences in spectra between low and high concentration samples were evident, suggesting the CTMs do not undergo a concentration-dependent oligomerisation. The predicted relative helical secondary structure for the CTMs was approximately 75%. An MRE_{222nm} of approximately -36000 deg.cm^2 dmol.res^{-1} was used as a benchmark for 100% alpha helical secondary structure [31]. Helical secondary structure was therefore quantified for the CTMs and a striking range of proportional helicities from 36% for C1 to 60% for C4 (Tab.1) was observed.

Heme B loading into the second tetrapyrrole binding site of C2 was previously shown to induce significant additional helical structure [8]. New bCTMs also exhibited the same phenomenon, with increases in relative alpha helical secondary structure of 10%, 13% and 12% for C1, C3 and C4

respectively. The most helical of all the new designs was C1C4 with a helical secondary structure content of 66%. Above approximately 50°C an increase in disordered secondary structure was observed from a significant positive shift in 222 nm ellipticity, indicative of denaturation. To probe this more thoroughly, thermal denaturation experiments were performed whilst monitoring the ellipticity at 222 nm from 5°C to 95°C. All designs exhibited sigmoidal melting trends with differing thermal denaturation midpoints (T_m) and variations in the degree of cooperativity evident. Interestingly, all new designs were significantly more thermostable than C2. The C3 CTM scaffold was the most thermostable with T_ms for the C3 CTM and bCTM of 59°C and 70°C respectively, a dramatic improvement on the thermal stability of C2 and BC2 proteins. bCTMs were considerably more thermostable and melted more cooperatively than the corresponding CTMs (Fig.5C,D).

To probe the CTM structures in more depth, the GROMACS molecular dynamics (MD) package [30] was used with the CHARMM forcefield [29] to produce in aqueous buffer at 27°C 1 μs, 570 ns and 300 ns simulations of CTMs, bCTMs and the C1C4 cCTM respectively. Root mean squared deviation (RMSD) traces confirmed that each CTM settled to a relatively stable conformational state by 1 μs (Fig.6.A-D). For C2 there is a sudden increase in RMSD of 0.2 nm at approximately 400 ns. Upon closer examination of the simulation helical fraying is evident at the C-terminus of the fourth helix that persists for the remainder of the simulation (Fig.6.F). This observation is unique to C2 and supports the CD data that suggests that C2 is the CTM most susceptible to thermal denaturation. Upon closer examination of the structural differences C1 and C3 or C2 and C4, it is clear that the differences in the orientation of heme within the same binding site indeed leads to a situation where the heme propionates are positioned either in the solvent (C3, C4) or within the hydrophobic protein core (C1, C2). By removing the charged propionates from the protein's hydrophobic interior, the former is clearly a more favorable conformation for protein stability. The charged groups of the C3 and C4 heme propionates were not only more solvent exposed in comparison to C1 and C2, but formed stabilising salt bridges with lysine residues from the hydrophilic helix regions. Furthermore, for bCTMs in which the orientation of heme B is not restricted by the CBM, heme B adopts an orientation whereby the heme propionates are oriented into bulk solvent and electrostatic interactions with surface lysines are formed.

There were other less significant perturbations in some of the structures: kinks in α-helices are observed particularly around the heme C binding sites, possibly reflecting the structural constraints afforded by covalent attachment of the heme C. Interestingly, in all designs part of the CBM adopted a relatively rigid, coil-like secondary structure and merged into loop regions, possibly accounting for some of the decrease in helical content relative to the hypothetical value for the CTM scaffold. This is

undoubtedly a result of placing the CBM near the N-termini of the α-helices and it may be possible to avoid this phenomenon in future designs by placing CBMs in central regions of the 4-helix bundles to enforce helical rigidity round the heme C binding site.

The simulations of bCTMs suggested some deviations from the canonical 4-helix bundle structure were also present. SiFig.5F shows BC2 at 1 μs and reveals an α-helix peeling away from the protein core, exposing hydrophobic core packing residues. The C-terminal end of helix 3 in BC4 (SiFig. 6H) bends perpendicularly away from the remainder of helix 3. Crucially however, all proteins were predominantly helical and in a 4-helix bundle topology (Fig6.E-H, SiFig.5E-H) for the duration of simulations and these perturbations only reflect individual 0.1 ns snapshots and hence are transient phenomena. Likewise, the C1C4 simulation revealed a predominantly helical protein with a 4-helix bundle topology consistent with most other diheme-containing bCTMs (SiFig.6).

Oxygen binding kinetics – Ferrous C2 was previously shown to bind molecular oxygen and form a stable oxyferrous species [8]. The UV-visible absorption properties of this intermediate were analogous to the neuroglobin oxyferrous species [25]. The manner of oxygen dissociation in C2 and neuroglobin were significantly different: C2 autoxidises and returns to the ferric state over two minutes at 15 ^{o}C, whereas oxygen dissociates from the ferrous heme iron of neuroglobin in low local concentrations of oxygen. This difference in oxygen release is most likely due to the low midpoint potential of the C2 heme relative to that of neuroglobin, favoring oxygen reduction over release. To probe the oxygen binding characteristics of the monoheme CTMs, ferrous CTMs were rapidly mixed with a saturating concentration of oxygen at 15°C in an stopped flow spectrophotometer housed in an anaerobic glovebox. Mixing with oxygenated buffer caused a rapid reduction in the intensity of the ferrous Q bands of all the CTMs and oxyferrous spectra identical to that of C2 subsequently formed in less than 1 second. These spectroscopic changes were observed for all new CTMs suggesting the proteins retained oxygen-binding properties despite the changes to heme C binding site. Following the formation of the oxyferrous complex, a slower autoxidative process was observed that ended with the formation of ferric CTM in approximately 100 seconds.

To examine oxygen binding and autoxidation kinetics more thoroughly, the absorbance at 569 nm was plotted against $\log_{10}$ time (Fig.7). This reveals two separate kinetic processes that occur between approximately 0-1 s and 1-100 s, attributed to oxygen binding and autoxidation, respectively. A large difference in the rates of each process allowed fitting to two single exponential models for the derivation of observed oxygen binding and autoxidation rates (Tab.1). C1 exhibited the fastest k_{obs} for oxygen binding of 19.3 ± 2.5 s^{-1} with the slowest oxygen binding kinetics observed for C3 with a k_{obs}

of 11.3 ± 2.2 s^{-1}. The slowest and fastest autoxidation rates of 0.25 ± 0.02 s^{-1} and 0.46 ± 0.04 s^{-1} were measured for C1 and C3 respectively, and are comparable to those of many oxygen activating cytochrome P450s [32].

Discussion

This work further illustrates not only the resilience of the maquette scaffold towards broad engineering strokes, particularly towards the placement of cofactor binding motifs that impose significant structural constraints on the protein, but also the promiscuous nature of the cytochrome *c* maturation apparatus of *E. coli*. With the exception of the diheme C maquettes C1C2 and C3C4, the CTMs and cCTMs described here are exported to the periplasm, recognized by the Ccm and post-translationally modified to graft heme C onto the protein backbone, regardless of recognition site positioning. It is well established that the *E. coli* Ccm exhibits promiscuous activity towards natural *c*-type cytochromes [33], efficiently processing prokaryotic and eukaryotic *c*-type cytochromes alike. With this work, we have demonstrated that this promiscuity extends to non-natural protein substrates, though the structural parameters influencing the efficiency of modification remain elusive. C3C4 suffers from issues relating to aggregation and therefore does not get processed by the Ccm; C2C3 and C1C4 are expressed, translocated and efficiently processed by the Ccm at both CBMs, though holo-C2C3 is prone to aggregation; C1C2 only achieves covalent incorporation of one heme C - most likely in the C1 position - indicating successful translocation and then recognition and maturation by the Ccm at CBM1 only. It is tantalizing to speculate that since CBM1 in C1C2 is situated at the N-terminus of the protein, CBM recognition and processing occurs as the peptide chain emerges from the Sec translocon. There is some evidence to suggest that maturation could occur in this manner; a natural cytochrome substrate with a non-cleavable Sec signal peptide was shown to undergo maturation while remaining attached to the translocon [34], revealing the accessibility of the Ccm machinery to Sec-anchored proteins. Whether the Ccm proteins naturally cluster near the Sec translocon remains to be investigated, thought it is clear that there are periplasmic proteins (e.g. the disulfide isomerase, DsbC) that contain a CXXCH motif yet avoid heme incorporation [35], thus arguing against a general mechanism involving concomitant translocation/maturation of the peptide chain. Of course, facile folding of a rapidly translocated peptide could kinetically trap the CXXCH motif within a folded protein or a folded protein subdomain whilst still attached to the translocon, thus rendering the motif inaccessible to the Ccm. An alternative mechanism exists whereby translocation occurs first, followed by apoprotein release and subsequent recognition/maturation by the Ccm [23,36]. The data presented here do not attempt to resolve this mechanistic question, though a greater understanding of this process would undoubtedly aid the design of future CTMs, particularly with regards to assembling more complex, multi-cofactor-containing maquettes.

Working under the hypothesis that apocytochrome Ccm substrates are released into the periplasm prior to maturation, we originally postulated that CTM thermal stability and rigidity might play a significant role in the efficiency of CBM recognition and modification, with more flexible proteins allowing better access to CBMs [8]. For C1C2 it is indeed a possibility that a well-folded, rigid structure is obtained after occupying CBM1, and it is this that kinetically inhibits covalent heme incorporation to CBM2 either when still attached to the translocon or released into the periplasm. However, since C1C4 undergoes highly efficient covalent heme incorporation to both CBMs, it is unlikely that a strong correlation exists between thermal stability and maturation, and the inability to modify certain CBMs in our protein might be due to an unfavorable local protein structure at the CBM rather than global protein stability. However, probing the exact non-natural substrate preferences of the Ccm is beyond the scope of this present study and will require an in-depth examination of both holo- and apo-CTM structural properties and folding kinetics.

The monoheme C CTMs are all expressed to high levels and processed efficiently through the Ccm, allowing the functional consequences of fixing heme orientation within the two original binding sites to be assessed. Our Molecular dynamics simulations indicate that the heme propionates of 2 variants (C1 and C2) point into the hydrophobic core while those of C3 and C4 are likely oriented into bulk solution. We would therefore predict that the relatively destabilizing act of placing negative charge in the CTM core would be reflected in the thermodynamic stability of C1 and C2, and we do indeed observe lower thermal melting points for C1 and C2 compared to the other variants. Despite these differences in thermal stability and heme orientation, it is striking to note that there are no large-scale functional differences with respect to heme redox potential, oxygen binding and autoxidation of the oxyferrous CTMs. However, what the changes in heme position and orientation have facilitated is a fine-tuning of these biophysical properties and the ability of the designer to select favorable attributes for future CTM designs.

Occupying the second tetrapyrrole binding sites of the CTMs with heme B serves to add rigidity to the maquette structures, in all cases increasing the percentage of observed helicity and significantly increasing the thermal stability of the scaffold. For C1, C2 and C3 this is associated with only minor changes in heme C redox potential, consistent with a lack of gross structural rearrangement, though C4 exhibits a significant positive (+44 mV) shift in heme C redox potential on binding heme B to the second site. Comparing the MD-derived structures obtained for C4 and BC4, we observe significant differences in the protein architecture, with a deviation away from a canonical 4-helix bundle structure when the second binding site is occupied. A rearrangement of side chains in the heme C binding site

is also evident, allowing for some rationalization of the heme C redox potential change based on the change in the local dielectric environment at the heme. Irrespective of the changes to the heme C redox potential, the bCTMs all exhibit redox potential differences of greater than 40 mV between hemes B and C, indicating that electron transfer from heme B to heme C would be significantly more favorable than in the reverse direction [37]. A similar, pronounced difference in heme C redox potentials (53 mV) is also observed for the C1C4 cCTM, thus highlighting the intrinsic asymmetry of the CTM scaffold and demonstrating that vectorial electron transfer is plausible in a maquette fully assembled and loaded with heme C cofactors *in vivo*. Since there are many natural multiheme *c*-type cytochromes known that engage in long distance intra- and interprotein electron transfer [38], it is not unreasonable to speculate that future multiheme CTMs could be integrated into natural multiprotein electron transfer pathways, such as the extracellular metal oxide reducing systems of the bacterial genus *Shewanella* [39], or hijack these complex systems to obtain electrons from promiscuous electron transfer hubs (e.g. CymA [40]) to perform a selected oxidoreductase activity *in vivo*. Indeed, rapid interprotein electron transfer transiently templated by electrostatic surface interactions has been demonstrated between heme-containing maquettes and natural proteins such as cytochrome *c* [7], thus highlighting the feasibility of integrating such components with cellular machinery to substitute, complement or hijack natural electron transfer components or pathways.

Conclusions

By successfully creating a suite of *in vivo*-assembled mono- and diheme CTMs, we have further demonstrated the flexibility and structural plasticity of the maquette scaffold, enabling the protein designer to select and fine-tune CTM properties and characteristics for future *de novo* oxidoreductase design. The simple modifications that we have described result in functional CTMs with thermal melting transitions spanning 35 ^{o}C, heme redox potentials over a 100 mV range, oxygen binding characteristics comparable to natural oxygen-activating enzymes and diheme proteins that exhibit heme redox potentials primed for vectorial intraprotein electron transfer. We anticipate that these CTM scaffolds will provide the starting point for more sophisticated maquettes that will ultimately enable functional oxidoreductase maquettes to be assembled and catalytically active within living organisms.

Acknowledgements:

The Authors wish to thank the BBSRC (grant no: BBI014063/1; DTG studentship to DWW), the Royal Society (through a University Research Fellowship to JLRA) and the European Research Council (partial studentship to DWW) for funding this research. We also wish to thank Prof. Dek Woolfson (Bristol) for kindly providing access to his equipment, Prof. Ian Collinson (Bristol) and his lab for

helpful discussions, and the University of Bristol's Advanced Computing Research Centre for access to the BlueCrystal supercomputer for running molecular dynamics simulations.

Bibliography

[1] I. Bertini, G. Cavallaro, A. Rosato, Cytochrome c: Occurrence and functions, Chem. Rev. 106 (2006) 90–115. doi:10.1021/cr050241v.

[2] B. Meunier, S.P. de Visser, S. Shaik, Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes, Chem. Rev. 104 (2004) 3947–3980. doi:10.1021/cr020443g.

[3] L.H. Xu, S. Fushinobu, S. Takamatsu, T. Wakagi, H. Ikeda, H. Shoun, Regio- and stereospecificity of filipin hydroxylation sites revealed by crystal structures of cytochrome P450 105P1 and 105D6 from Streptomyces avermitilis, J. Biol. Chem. 285 (2010) 16844–16853. doi:10.1074/jbc.M109.092460.

[4] S. Takahashi, Y.S. Yeo, Y. Zhao, P.E. O'Maille, B.T. Greenhagen, J.P. Noel, et al., Functional Characterization of Premnaspirodiene Oxygenase, a Cytochrome P450 Catalyzing Regio- and Stereo-specific Hydroxylations of Diverse Sesquiterpene Substrates, J. Biol. Chem. 282 (2007) 31744–31754. doi:10.1074/jbc.M703378200.

[5] C.T. Armstrong, D.W. Watkins, J.L.R. Anderson, Constructing manmade enzymes for oxygen activation., Dalton Trans. 42 (2013) 3136–50. doi:10.1039/c2dt32010j.

[6] R.L. Koder, J.L.R. Anderson, L.A. Solomon, K.S. Reddy, C.C. Moser, P.L. Dutton, Design and engineering of an O(2) transport protein., Nature. 458 (2009) 305–309. doi:10.1038/nature07841.

[7] T.A. Farid, G. Kodali, L.A. Solomon, B.R. Lichtenstein, M.M. Sheehan, B.A. Fry, et al., Elementary tetrahelical protein design for diverse oxidoreductase functions., Nat. Chem. Biol. 9 (2013) 826–833. doi:10.1038/nchembio.1362.

[8] J.L.R. Anderson, C.T. Armstrong, G. Kodali, B.R. Lichtenstein, D.W. Watkins, J. Mancini, et al., Constructing a man-made, oxygen binding c-type cytochrome maquette in vivo., Chem. Sci. (2013). doi:10.1039/c3sc52019f.

[9] F. Hollmann, I.W.C.E. Arends, K. Buehler, A. Schallmey, B. Bühler, Enzyme-mediated oxidations for the chemist, Green Chem. 13 (2011) 226. doi:10.1039/c0gc00595a.

[10] J.C. Lewis, P.S. Coelho, F.H. Arnold, Enzymatic functionalization of carbon-hydrogen bonds., Chem. Soc. Rev. 40 (2011) 2003–2021. doi:10.1039/c0cs00067a.

[11] B.R. Lichtenstein, T.A. Farid, G. Kodali, L.A. Solomon, J.L.R. Anderson, M.M. Sheehan, et al., Engineering oxidoreductases: maquette proteins designed from scratch, Biochem. Soc. Trans. 40 (2012) 561–566. doi:10.1042/BST20120067.

[12] B.R. Gibney, F. Rabanal, J.J. Skalicky, A.J. Wand, P.L. Dutton, Design of a unique protein scaffold for maquettes, J. Am. Chem. Soc. 119 (1997) 2323–2324. doi:10.1021/ja963561s.

[13] B. Discher, R.L. Koder, C.C. Moser, P.L. Dutton, Hydrophilic to amphiphilic design in redox protein maquettes, Curr. Opin. Chem. Biol. 7 (2003) 741–748. doi:10.1016/j.cbpa.2003.10.013.

[14] D.W. Watkins, C.T. Armstrong, J.L.R. Anderson, De novo protein components for oxidoreductase assembly and

biological integration., Curr. Opin. Chem. Biol. 19 (2014) 90–8. doi:10.1016/j.cbpa.2014.01.016.

[15] B.R. Gibney, Y. Isogai, F. Rabanal, K.S. Reddy, A.M. Grosset, C.C. Moser, et al., Self-assembly of heme A and heme B in a designed four-helix bundle: implications for a cytochrome c oxidase maquette., Biochemistry. 39 (2000) 11041–11049. doi:10.1021/bi000925r.

[16] M.R. Razeghifard, T. Wydrzynski, Binding of Zn-chlorin to a synthetic four-helix bundle peptide through histidine ligation., Biochemistry. 42 (2003) 1024–1030. doi:10.1021/bi026787i.

[17] R.E. Sharp, C.C. Moser, F. Rabanal, P.L. Dutton, Design, synthesis, and characterization of a photoactivatable flavocytochrome molecular maquette., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 10465–70. doi:10.1073/pnas.95.18.10465.

[18] F. Rabanal, B. Gibney, W. DeGrado, Engineering photosynthesis : synthetic redox proteins, Inorganica Chim. Acta. 243 (1996) 213–218. doi:10.1016/0020-1693(95)04910-X.

[19] J.M. Shifman, C.C. Moser, W.A. Kalsbeck, D.F. Bocian, P.L. Dutton, Functionalized de novo designed proteins: mechanism of proton coupling to oxidation/reduction in heme protein maquettes., Biochemistry. 37 (1998) 16815–27. doi:10.1021/bi9816857.

[20] L.A. Solomon, G. Kodali, C.C. Moser, P.L. Dutton, Engineering the assembly of heme cofactors in man-made proteins, J. Am. Chem. Soc. 136 (2014) 3192–3199. doi:10.1021/ja411845f.

[21] A.M. Grosset, B.R. Gibney, F. Rabanal, C.C. Moser, P.L. Dutton, Proof of principle in a de novo designed protein maquette: an allosterically regulated, charge-activated conformational switch in a tetra-alpha-helix bundle., Biochemistry. 40 (2001) 5474–5487. doi:10.1021/bi002504f.

[22] C. Hunter, E. Lloyd, L. Eltis, S. Rafferty, Role of the heme propionates in the interaction of heme with apomyoglobin and apocytochrome b5, Biochemistry. 36 (1997) 1010–7. doi:10.1021/bi961385u.

[23] J.M. Stevens, D.A.I. Mavridou, R. Hamer, P. Kritsiligkou, A.D. Goddard, S.J. Ferguson, Cytochrome c biogenesis System I., FEBS J. 278 (2011) 4170–4178. doi:10.1111/j.1742-4658.2011.08376.x.

[24] J.W.A. Allen, N. Leach, S.J. Ferguson, The histidine of the c-type cytochrome CXXCH haem-binding motif is essential for haem attachment by the Escherichia coli cytochrome c maturation (Ccm) apparatus., Biochem. J. 389 (2005) 587–592. doi:10.1042/BJ20041894.

[25] S. Dewilde, L. Kiger, T. Burmester, T. Hankeln, V. Baudin-Creuza, T. Aerts, et al., Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family., J. Biol. Chem. 276 (2001) 38949–38955. doi:10.1074/jbc.M106438200.

[26] J. Chiu, P.E. March, R. Lee, D. Tillett, Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h., Nucleic Acids Res. 32 (2004) e174. doi:10.1093/nar/gnh172.

[27] T. Uchida, J.M. Stevens, O. Daltrop, E.M. Harvat, L. Hong, S.J. Ferguson, et al., The interaction of covalently bound heme with the cytochrome c maturation protein CcmE., J. Biol. Chem. 279 (2004) 51981–8. doi:10.1074/jbc.M408963200.

[28] E.A. Berry, B.L. Trumpower, Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra., Anal. Biochem. 161 (1987) 1–15. doi:10.1016/0003-2697(87)90643-9.

[29] E. Lindahl, P. Bjelkmar, P. Larsson, M.A. Cuendet, B. Hess, Implementation of the charmm force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models, J. Chem. Theory Comput. 6 (2010) 459–466. doi:10.1021/ct900549r.

[30] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56. doi:10.1016/0010-4655(95)00042-E.

[31] E.K. O'Shea, R. Rutkowski, P.S. Kim, Evidence that the leucine zipper is a coiled coil., Science. 243 (1989) 538–542. doi:10.1126/science.2911757.

[32] L.G. Denisov, Y. V Grinkova, M.A. McLean, S.G. Sligar, The one-electron autoxidation of human cytochrome P450 3A4, J. Biol. Chem. 282 (2007) 26865–26873. doi:10.1074/jbc.M704747200.

[33] J.W.A. Allen, S.J. Ferguson, What is the substrate specificity of the System I cytochrome c biogenesis apparatus?, Biochem. Soc. Trans. 34 (2006) 150–1. doi:10.1042/BST0340150.

[34] L. Thöny-Meyer, P. Künzler, Translocation to the periplasm and signal sequence cleavage of preapocytochrome c depend on sec and lep, but not on the ccm gene products., Eur. J. Biochem. 246 (1997) 794–799.

[35] D.A.I. Mavridou, M. Braun, L. Thöny-Meyer, J.M. Stevens, S.J. Ferguson, Avoidance of the cytochrome c biogenesis system by periplasmic CXXCH motifs., Biochem. Soc. Trans. 36 (2008) 1124–8. doi:10.1042/BST0361124.

[36] D.A.I. Mavridou, J.M. Stevens, L. Mönkemeyer, O. Daltrop, K. Di Gleria, B.M. Kessler, et al., A pivotal heme-transfer reaction intermediate in cytochrome c biogenesis, J. Biol. Chem. 287 (2012) 2342–2352. doi:10.1074/jbc.M111.313692.

[37] C.C. Page, C.C. Moser, X. Chen, P.L. Dutton, Natural engineering principles of electron tunnelling in biological oxidation-reduction., Nature. 402 (1999) 47–52. doi:10.1038/46972.

[38] T.A. Clarke, M.J. Edwards, A.J. Gates, A. Hall, G.F. White, J. Bradley, et al., Structure of a bacterial cell surface decaheme electron conduit., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 9384–9389. doi:10.1073/pnas.1017200108.

[39] N.S. Wigginton, K.M. Rosso, B.H. Lower, L. Shi, M.F. Hochella, Electron tunneling properties of outer-membrane decaheme cytochromes from Shewanella oneidensis, Geochim. Cosmochim. Acta. 71 (2007) 543–555. doi:10.1016/j.gca.2006.10.002.

[40] J.M. Myers, C.R. Myers, Role of the tetraheme cytochrome CymA in anaerobic electron transport in cells of Shewanella putrefaciens MR-1 with normal levels of menaquinone, J. Bacteriol. 182 (2000) 67–75. doi:10.1128/JB.182.1.67-75.2000.

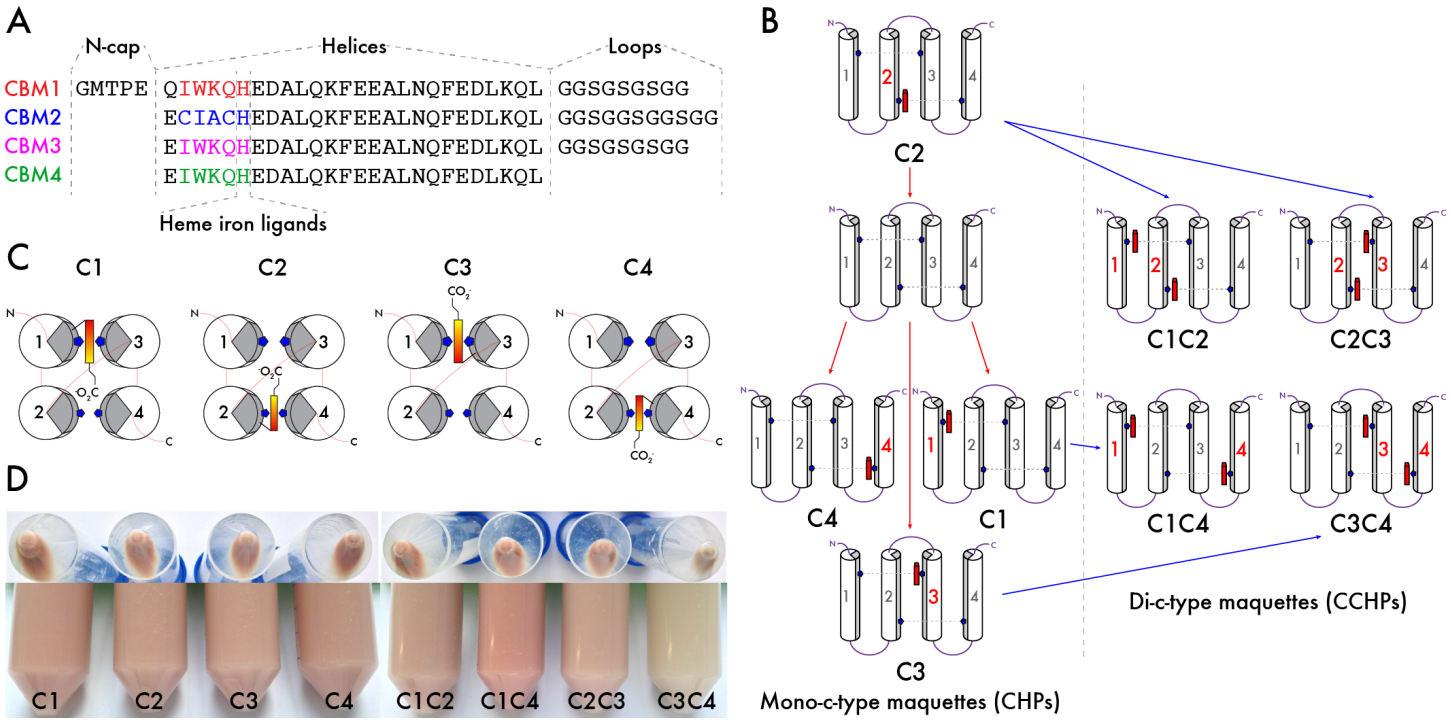

Figure 1: Sequence, design and expression of CTMs and cCTMs. A - Protein sequence of C2. **B** – Design schematic of the CTMs and cCTMs. Cartoon CTMs illustrate how the new designs were achieved by successive rounds of mutagenesis to C2. Histidine side chains (blue pentagons), hemes C (red/yellow rectangles), loops (red lines) and helices (cylinders) with hydrophobic (grey) and hydrophilic (white) regions are shown. **C** - Bottom up cartoon representation of mono-heme C maquettes, emphasising the orientation of heme C (red/yellow cuboid) when tethered to CBM1 or CBM3 and CBM2 or CBM4 by thioether bonds (black line). **D** – Centrifuged (top) and resuspended (bottom) *E. coli* cells after four hours of CTM expression.

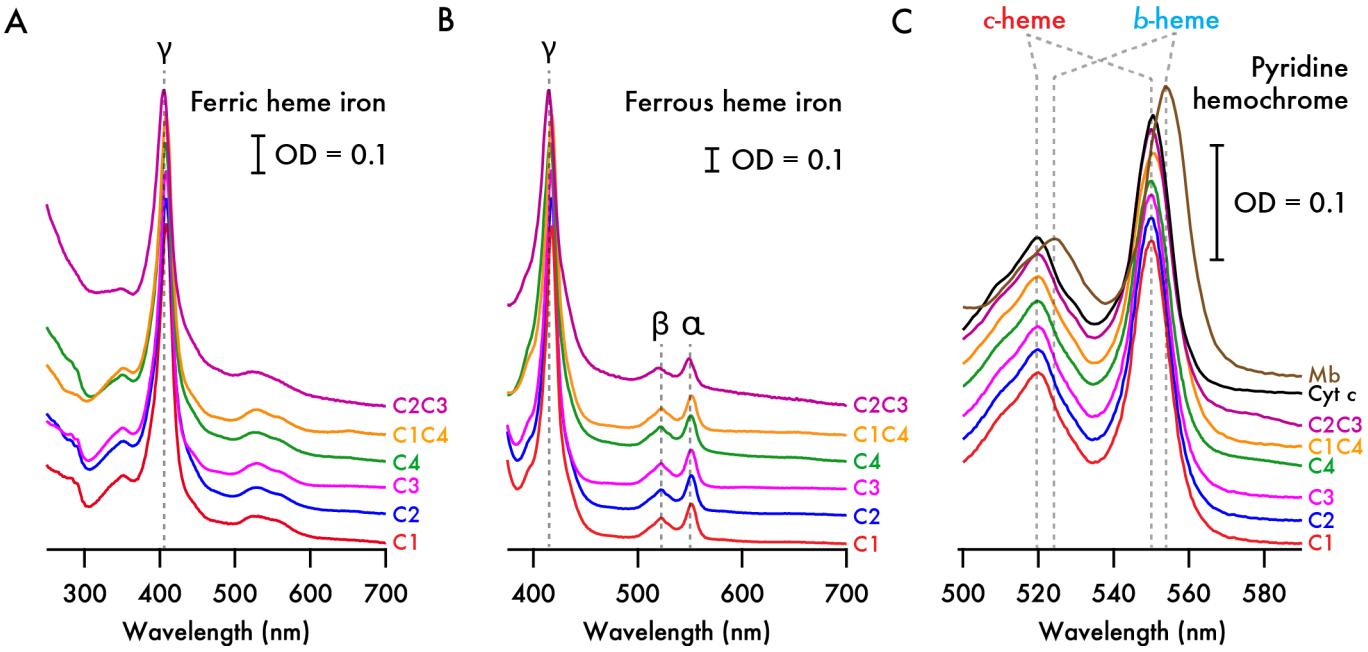

Figure 2: UV-visible absorption properties of CTMs and cCTMs. UV-visible absorption spectra of 5 µM ferric CTM and cCTM in 100 mM KCl, 20 mM K phosphate, pH 7.5 (offset on the Y axis) (**A**). Thereafter samples were reduced with excess sodium dithionite and ferrous spectra recorded (**B**). 5 µM pyridine hemochrome samples (**C**) were prepared as described in Berry *et. al.* [28] to confirm the association of heme C to maquettes. Myoglobin (heme B) and cytochrome *c* (heme C) pyridine hemochrome spectra are shown as controls.

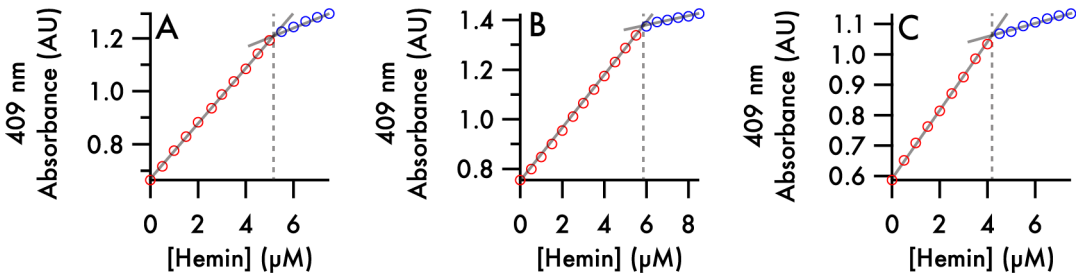

Figure 3: Heme titrations with C1, C3 and C4. Heme B binding stoichiometries of C1 (**A**), C3 (**B**) and C4 (**C**) were assessed by titrating CTMs with hemin in DMSO (0.5 mM). UV-visible spectra were recorded after 0.5 μM aliquots of hemin were added to 5 μM CTM solution in 100 mM KCl, 20 mM K phosphate, pH 7.5. Absorbance at 409 nm was plotted against added hemin concentration. Data for pre-saturation (red circles) and post-saturation (blue circles) of heme are fitted to separate linear models (grey lines), revealing the binding stoichiometry of heme B to the second tetrapyrrole binding site on the CTMs.

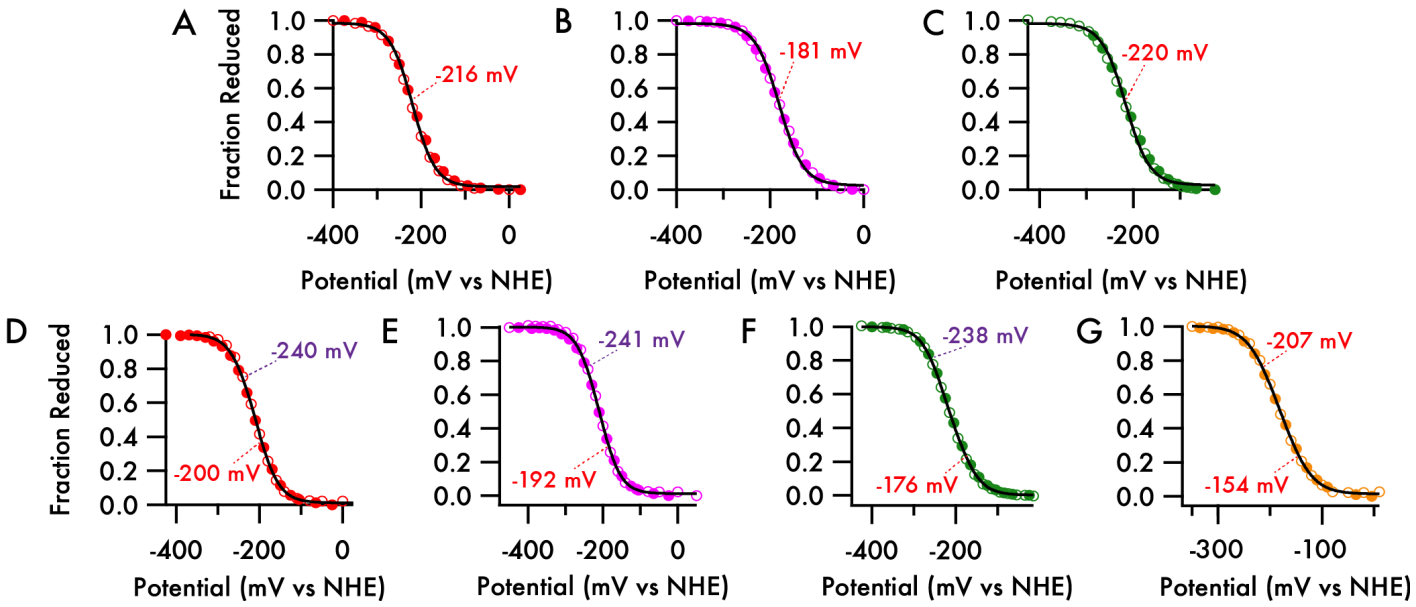

Figure 4: Redox potentiometry of CTMs, bCTMs and C1C4. The redox electrochemical properties of 50-100 μM C1/BC1 (**A**/**D**), C3/BC3 (**B**/**E**), C4/BC4 (**C**/**F**) and C1C4 (**G**) were analysed with OTTLE potentiometry. A series of reducing (open circles) then oxidising potentials (filled circles) were applied to samples, measuring the UV-visible spectrum after 30 minutes of equilibration. Ferrous soret absorbance (418 nm for CTMs and C1C4, 424 nm for bCTMs) was normalised to fraction of reduced protein, plotted against potential and fitted to a Nernst model (black lines) adapted to accommodate the quantity of hemes per protein. *C*- (red text labels) and *b*- (purple text labels) heme iron E_ms are shown. Potentials are quoted versus the Nernst Hydrogen Electrode (NHE).

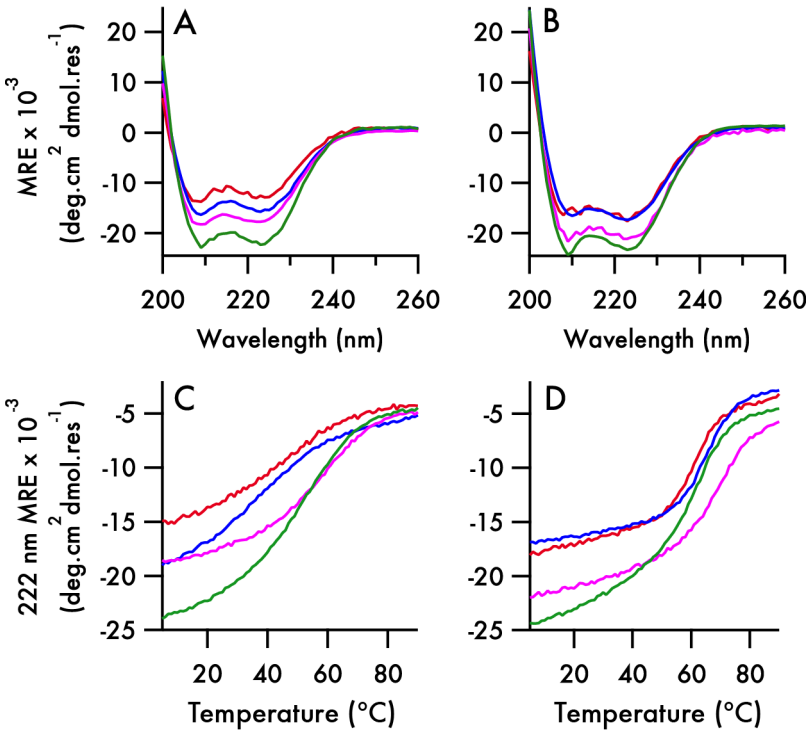

Figure 5: Far-UV circular dichroism spectra and 222 nm thermal denaturation traces of CTMs and bCTMs. Far-UV CD spectra of 10 µM C1/BC1 (red), C2/BC2 (blue), C3/BC3 (pink) and C4/BC4 (green) in 100 mM KCl, 20 mM K phosphate, pH 7.5 were measured at 25°C for CTMs (**A**) and bCTMs (**B**). Thermal denaturation traces were recorded by monitoring 222 nm ellipticity with 1°C intervals from 5°C to 95°C for CTMs (**C**) and bCTMs (**D**). Y-axis units were converted to mean residue ellipticity by normalising ellipticity to peptide bond concentration and cuvette pathlength.

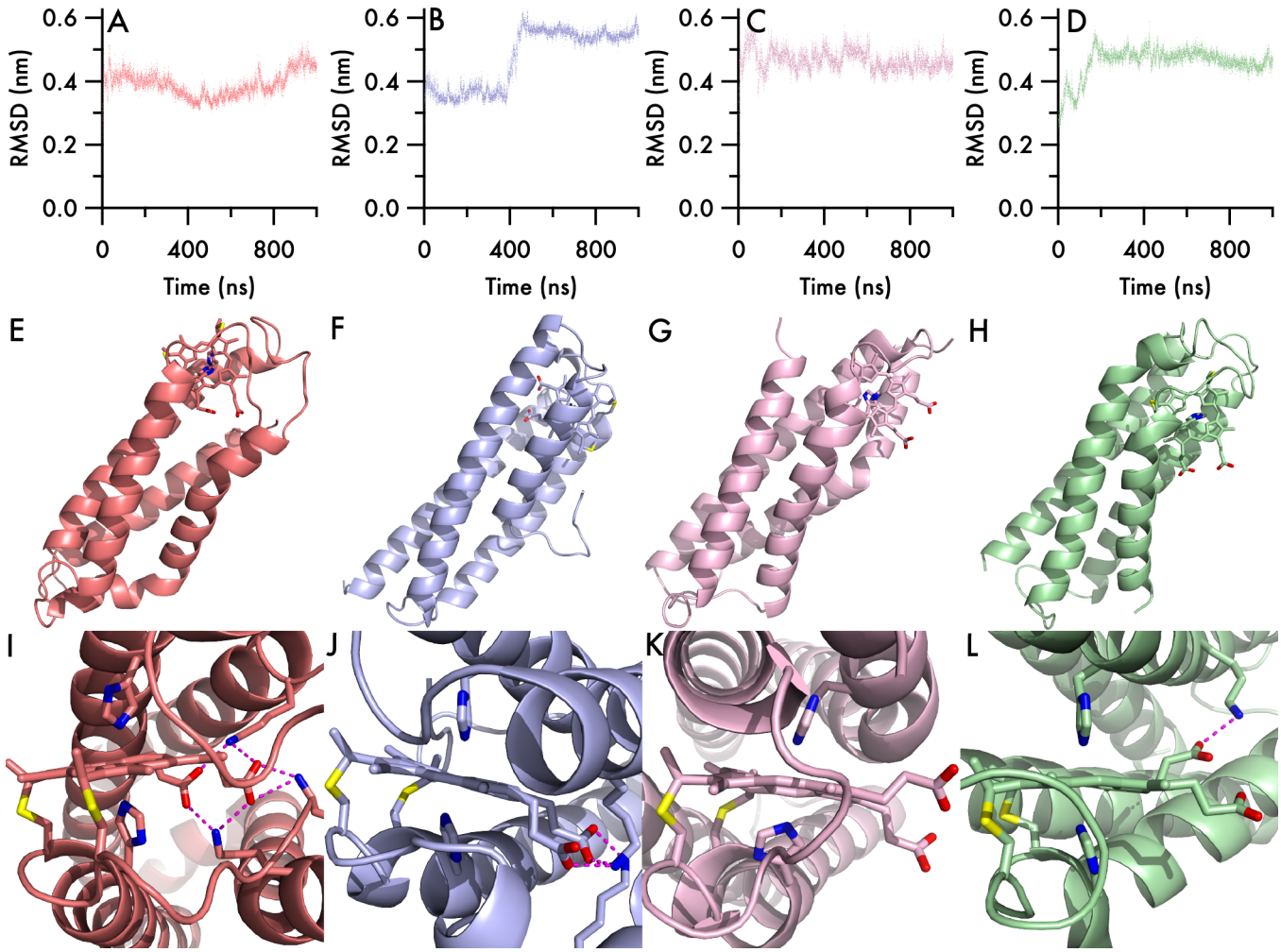

Figure 6: Molecular dynamics simulations of CTMs. 1 μs MD simulations of C1 (red), C2 (blue), C3 (pink) and C4 (green). Root mean squared deviation traces (**A**-**D**) were derived from the deviation in alpha carbon coordinates with respect to the starting structure calculated by GROMACS [29]. A side view (**E**-**H**) and close up view of heme C (**I**-**L**) for CTMs after 1 μs are shown with alpha helices represented as ribbons, and heme as sticks. Thioether bond sulfur atoms (yellow), nitrogen atoms (blue) and heme propionate oxygens (red) are also highlighted. Dashed magenta lines highlight salt bridges between heme propionates and lysine side chains.

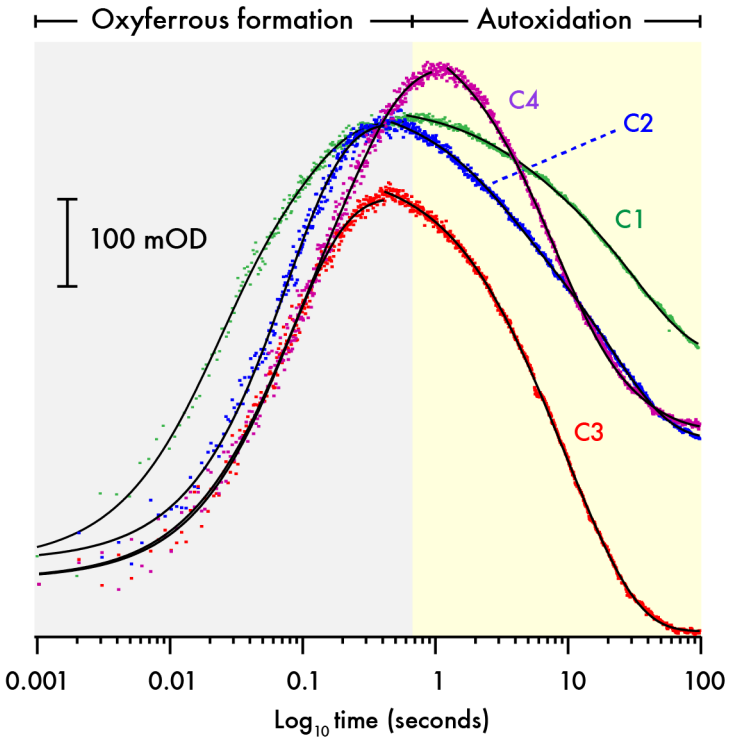

Figure 7: Oxygen binding kinetics for CTMs. 5 μM C1 (**A**), C2 (**B**), C3 (**C**) and C4 (**D**) were mixed with 540 μM oxygen in an anaerobic stopped-flow spectrophotometer. The absorbance at 569 nm was plotted vs $\log_{10}$ time (colored dots). Two separate processes were observed, an initial fast transition (over approximately 0-0.7 seconds) corresponding to oxygen binding and a second slower transition (over approximately 0.7-100 seconds) corresponding to autoxidation. Each process was fitted to a single exponential function (black traces) for the derivation of the observed rate constants (k_{obs}) for oxygen binding and autoxidation (Tab.1).

Property	C1/BC1	C2/BC2	C3/BC3	C4/BC4	C1C2	C1C4	C2C3	C3C4
Fully processed by Ccm	Y	Y	Y	Y	N	Y	Y	N
Soluble and monomeric	Y	Y	Y	Y		Y	N	
Heme binding stoichiometry	1:1	1:1	1:1	1:1		NA		
% Helicity	36 / 46	44 / 47	49 / 58	60 / 62		66		
T_m (°C)	47 / 61	35 / 65	59 / 70	51 / 62		60		
Heme C E_m (mV vs NHE) (CTMs and cCTMs only)	-216 ± 1	-198 ± 1	-181 ± 1	-220 ± 1		-207 ± 1 / -154 ± 1		
Heme B/heme C E_m (mV vs NHE) (bCTMs only)	-240 ± 4 / -200 ± 3	-250 ± 6 / -192 ± 4	-241 ± 10 / -192 ± 7	-238 ± 2 / -176 ± 3		NA		
k_{obs}(O_2 on) (s^{-1})	19.3 ± 2.5	11.7 ± 1.7	11.3 ± 2.2	15.3 ± 2.8		NA		
k_{obs}(Autoxidation) (s^{-1})	0.25 ± 0.02	0.27 ± 0.01	0.46 ± 0.04	0.27 ± 0.02		NA		

Table 1: Summary of CTM, cCTM and bCTM expression and biophysical properties.